

WHAT IS CLAIMED IS:

1. An audio player, comprising:  
a processor;  
a memory connected to said processor;  
5 an audio data buffer allocated in said memory to store first and second audio data received from an audio server in response to an audio selection signal, said first audio data representing a first portion of an audio clip, said second audio data representing a second portion of said audio clip, said first audio data received before said second audio data;  
10 a metadata buffer allocated in said memory to store metadata received following said audio selection signal;  
an audio transducer connected to said processor; and  
15 playback software running on said processor, said playback software converting said first audio data into analog audio data, said audio transducer generating sound from said analog audio data before said receipt of said second audio data, said playback software converting said metadata into a visual format.
2. The audio player described in Claim 1, wherein said audio data buffer is dynamically resized based on a rate at which said first audio data is received from said audio server.
- 20 3. The audio player described in Claim 1, wherein first content in said metadata and second content in said audio clip correlate said metadata and said audio clip.
4. The audio player described in Claim 1, further comprising  
a display connected to said processor, wherein said playback software is  
25 configured to present on said display said visually formatted metadata.
5. The audio player described in Claim 4, wherein said playback software delays said presentation of said metadata following its receipt.
6. The audio player described in Claim 5, wherein said visually formatted metadata represents text.
- 30 7. The audio player described in Claim 5, wherein said visually formatted metadata represents an image.

8. The audio player described in Claim 4, wherein said playback software presents said visually formatted metadata before said audio transducer generates said sound.

5 9. The audio player described in Claim 8, wherein said visually formatted metadata represents text.

10 11. The audio player described in Claim 8, wherein said visually formatted metadata represents an image.

12. The audio player described in Claim 4, further comprising:  
timing data stored in said memory, said playback software commencing  
said presentation of said visually formatted metadata in accordance with said  
timing data.

13. A method of playing audio data, the method comprising:  
selecting an audio source;  
receiving in a first buffer in a computer-readable memory first audio data  
representing a portion of audio information from said audio source;  
receiving in said first buffer second audio data representing a second  
portion of said audio information from said audio source, said second audio data  
received after said receipt of said first audio data;  
receiving metadata in a second buffer;  
generating sound from said first audio data before said receipt of said  
second audio data; and  
displaying a visual representation of said metadata.

14. The method described in Claim 12, wherein first content in said metadata  
and second content in said audio information correlate said metadata and said audio  
information.

15. The method described in Claim 12, wherein said first and second audio  
data are received via a data communication network.

16. The method described in Claim 12, wherein said visual representation of  
said metadata includes an image.

17. The method described in Claim 12, wherein said visual representation of  
said metadata includes a text character.

17. The method described in Claim 12, wherein said displaying occurs before said generating said sound.

18. The method described in Claim 12, wherein said displaying is delayed until after said generating said sound.

5 19. The method described in Claim 12, wherein said displaying occurs during said generating said sound.

20. An aggregation of media data, comprising:

first media data in a first memory area, said first media data representing a first portion of an audio clip received from a remote audio center, said first media data used to generate sound prior to a time  $t_1$ ;

10

second media data in a second memory area, said second media data representing a second portion of said audio clip, said second media data received from said remote audio center after said time  $t_1$ ;

third media data in a third memory area, said third media data used to generate a visual display before said time  $t_1$ ; and

15

said first, second and third media data received in response to an audio selection signal.

21. The aggregation of media data described in Claim 20, further comprising:

fourth media data representing a third portion of said audio clip, said fourth media data received from said remote audio center after said receipt of said second media data, said fourth media data at least partly stored in said first memory area.

20

25